

Public Health Data Resources

June 2014

Doug Joubert, NIH Library

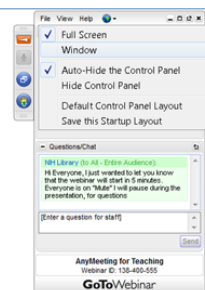
NIH Library | <http://nihlibrary.nih.gov>

Office of Research Services
National Institutes of Health
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Office of Research Services
Serving the NIH Community

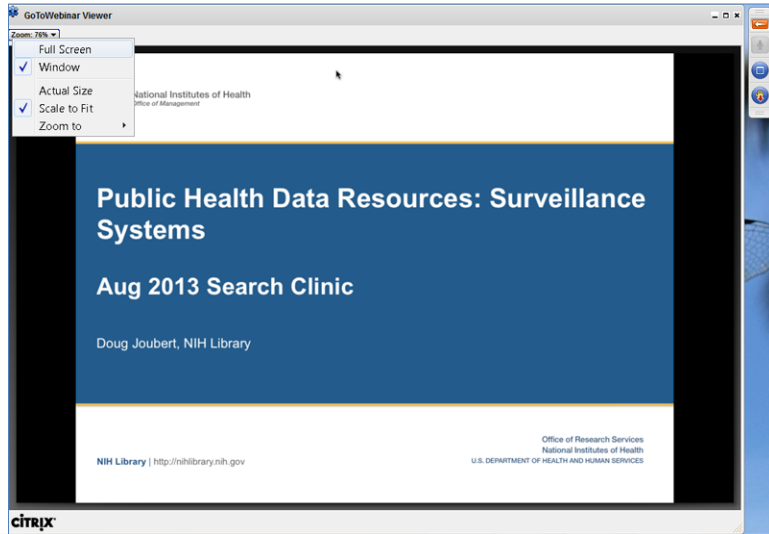
Webinar: Working with Views

Public Health and GIS Informationist for the ONC



Office of Research Services
Serving the NIH Community

Webinar: Resizing your screen



The screenshot shows a GoToWebinar Viewer window. The top menu bar includes 'Zoom: 75%', 'Full Screen', 'Window', 'Actual Size', 'Scale to Fit', and 'Zoom to'. The main content area displays a presentation slide from the National Institutes of Health (NIH) Office of Research Services. The slide title is 'Public Health Data Resources: Surveillance Systems', dated 'Aug 2013 Search Clinic', presented by 'Doug Joubert, NIH Library'. The slide footer includes the NIH Library logo and the URL 'http://nihlibrary.nih.gov'. The bottom of the window shows the Citrix logo.

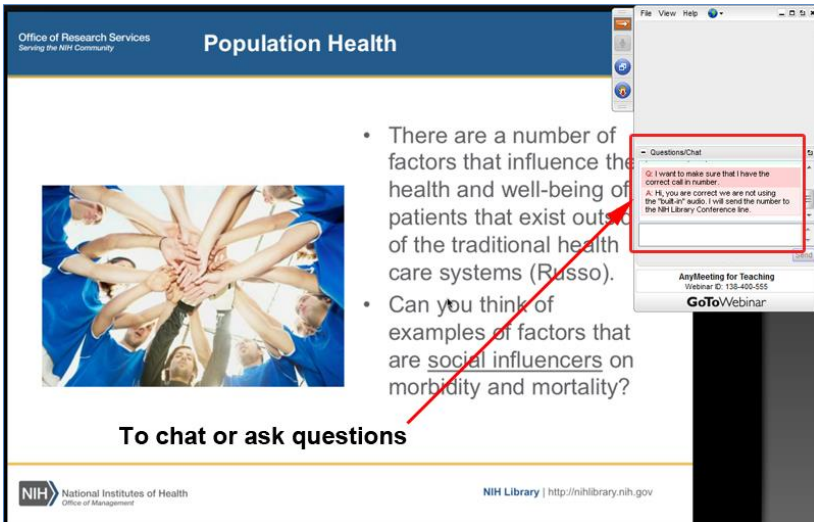
NIH National Institutes of Health
Office of Management

NIH Library | <http://nihlibrary.nih.gov>

NIH Library | <http://nihlibrary.nih.gov>

Office of Research Services
Serving the NIH Community

Webinar: Chating/Questions



The screenshot shows a GoToWebinar Viewer window. The top menu bar includes 'File', 'View', and 'Help'. The main content area displays a presentation slide from the NIH Office of Research Services titled 'Population Health'. The slide features a photograph of a group of people in blue shirts huddled together, and a bulleted list of questions. A red box highlights the 'Questions/Chat' window on the right side of the viewer, which contains a message from a participant asking for a correct call number. The bottom of the window shows the NIH logo and the URL 'http://nihlibrary.nih.gov'.

Office of Research Services
Serving the NIH Community

Population Health

- There are a number of factors that influence the health and well-being of patients that exist outside of the traditional health care systems (Russo).
- Can you think of examples of factors that are social influencers on morbidity and mortality?

To chat or ask questions

NIH National Institutes of Health
Office of Management

NIH Library | <http://nihlibrary.nih.gov>

NIH Library | <http://nihlibrary.nih.gov>

Outline

- Genesis of this class.
- Introduction to population health.
- Overview of the common data types in public health research.
- Basic statistical and epidemiology terms used in public health.
- How public health data is collected.
- Data focus: public health surveillance.
- Hands-on exercise.

Online companion to this series

Health Data Resources Tags: health data, public health

This online guide contains information on finding, using, and analyzing health and population data. The focus of this guide is on publicly available data sources and freely available analysis tools.

Last Updated: Sep 14, 2013 | URL: <http://nihlibrary.campusguides.com/hdata> | [Print Guide](#) | [RSS Updates](#) | [Email Alerts](#) | [SHARE](#) [f](#) [t](#) [in](#)

[Home](#) | [Common Data Types in Public Health Research](#) | [Demographic Data](#) | [Disease Registries](#) | [Financial or Service Utilization Data](#)
[Hospital Data](#) | [Surveillance Data](#) | [Survey Data](#) | [Guide Sources](#)

Home | [Comments\(0\)](#) | [Print Page](#) | Search: | This Guide [v](#) | Search

Introduction

This online guide contains information on finding, using, and analyzing health and population data. The guide was developed as an online companion to the "Keeping up with Public Health Series," class taught by Alicia Livinski, Doug Joubert, and Nancy Terry.

If you need a one-on-one consultation on data sources, data visualization and management, or basic data analysis please contact Doug Joubert at the NIH Library.

Comments (0)


Getting Help and Training

NIH Librarians and Informationist are available to create custom training solutions for your office, lab, or branch. Please consult our current training schedule, or request a tutorial. Currently the following public health webinars are being offered by NIH Library staff.

If you need a one-on-one consultation on data sources, data visualization and management, or basic data analysis please contact Doug Joubert at the NIH Library.

Comments (0)

Biomedical Librarian | Informationist



Douglas Joubert

Chat is offline [f](#) [t](#) [in](#)

<http://nihlibrary.campusguides.com/hdata>

Population Health

- The population health model seeks to understand the causes of the systematic differences in the health between different groups (Kindig & Stoddart, 2003).
- Population health includes both **health outcomes** and patterns of **health determinants**, and the **policies and interventions** that link these two (Kindig & Stoddart, 2003).



Social Gradients of Health

- The social gradient in health refers to the fact that inequalities in population health status are related to inequalities in social status (Kindig & Stoddart, 2003; Kosteniuk & Dickinson, 2003; World Health Organization, 2013).



Social Indicators of Health



- Factors that influence the health and well-being of patients that exist outside of the traditional health care systems (Russo, 2011).
- Can you think of examples of factors that are social influencers on morbidity and mortality?

Determinants of Health

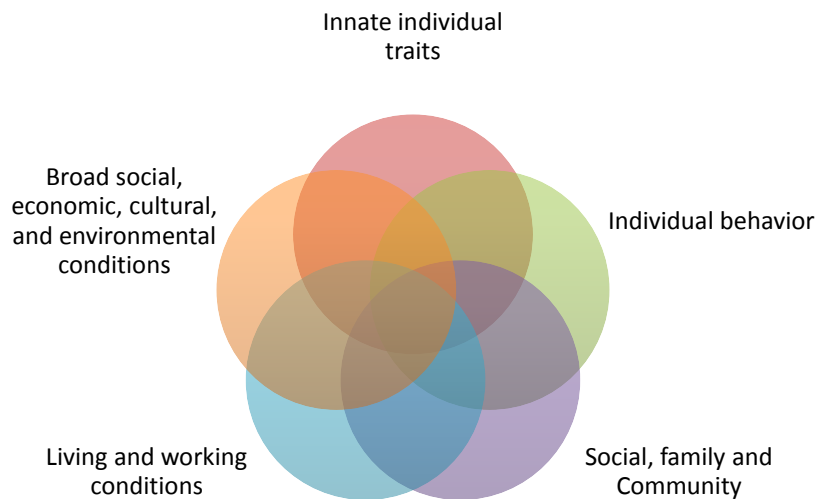
- Population health scientists use the term ***determinants of health***, rather than the ***cause*** or ***factor*** (Russo, 2011).
- Determinants of health can be divided into the following categories (Kindig & Stoddart, 2003; Kosteniuk & Dickinson, 2003; World Health Organization, 2013):
 - Social and economic
 - Physical environment
 - Genetics
 - Medical care
 - Health-related behaviors

Health Disparities

- A fundamental element in population health is health disparities.
- Health disparities are the differences in health status among groups.
- These differences can affect how frequently a disease affects a group, how many people get sick, or how often the disease causes death (Adler 1994).



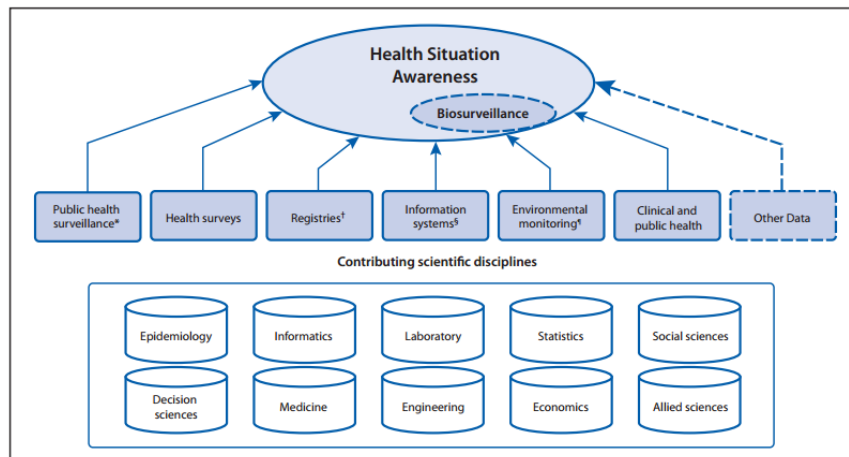
Dimensions of Population Health



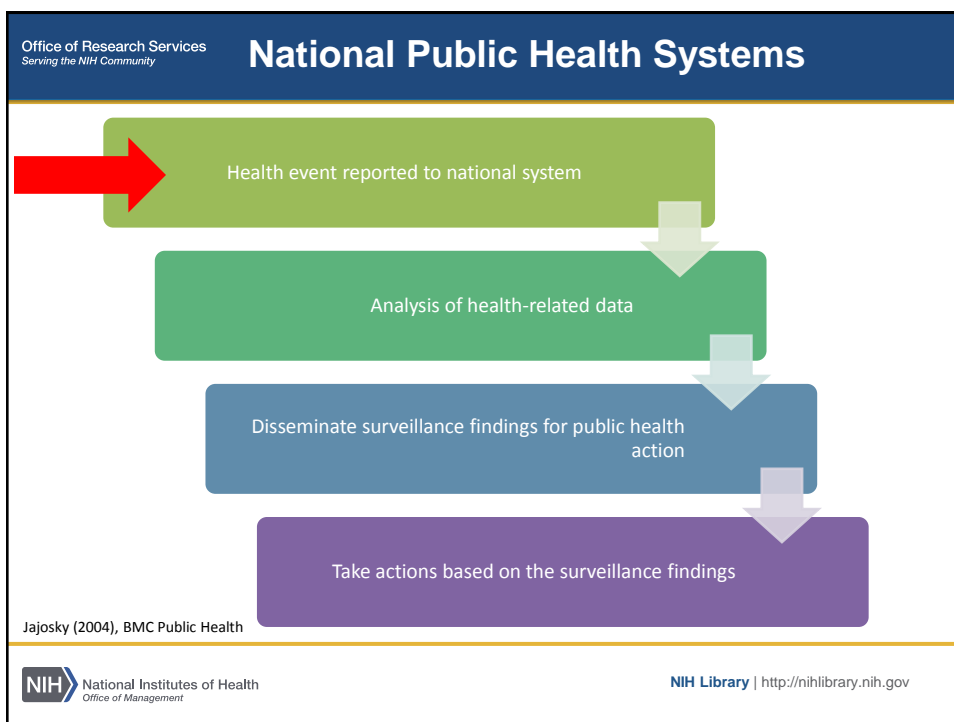
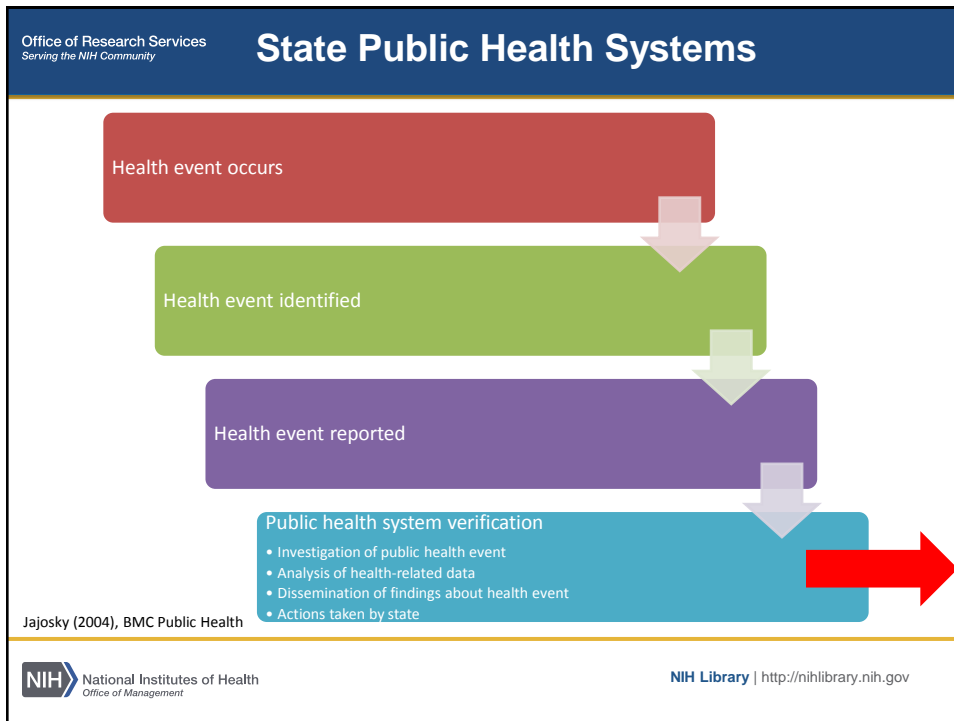
How public health data is collected



Health Situation Awareness



(Thacker, Qualters, & Lee, 2012).



Disease Surveillance

- Disease surveillance: monitoring distributions and trends in morbidity (affected by a disease) and mortality data collected for specific populations and geographical areas.
- Notifiable disease reports and vital records are the two health data sources available at the local levels in all states.

(Cromley & McLafferty, 2012b).

National Notifiable Diseases Surveillance Systems

- The National Notifiable Diseases Surveillance Systems (NNDSS) is operated by the CDC and the Council of State and Territorial Epidemiologists (CTSE).
- Reporting to the national system is voluntary, and states generally only report internationally quarantinable diseases in compliance with WHO International Regulations.
- There are approximately 50 infectious diseases designated as notifiable at the national level.
- This list changes periodically, and reporting practices may differ between states.

(Cromley & McLafferty, 2012b).

Data types in public health research

Quantitative Data

- Quantitative data is measurable, often used for comparisons, and involves counting of people, behaviors, conditions, or other discrete events.
- Quantitative data uses numbers to determine the what, who, when, and where of health-related events.
- Examples of quantitative data include: age, weight, temperature, or the number of people suffering from diabetes.

Wang, 2013.

Qualitative Data

- Qualitative data uses words to describe a particular health-related event.
- This data can be observed, but not measured.
- Involves observing people in selected places and listening to discover how they feel and why they might feel that way.
- Examples of qualitative data include: male/female, smoker/non-smoker, or questionnaire response (agree, disagree, neutral).

Wang, 2013.
Romano.

Primary Data Sources

- Primary data is data that **you** or **your** colleagues collect specifically for the purpose of answering **your** research question.
- Advantages to Using Primary Data
 - Collect exactly the data elements that you need to answer your research question.
 - Can test an intervention, such as an experimental drug or an educational program, in the purest way (a double-blind randomized controlled trial).
 - Control the data collection process, so you can ensure data quality, minimize the number of missing values, and assess the reliability of your instruments.

Romano.

Secondary Data Sources

- Existing data collected for another purpose, that you use to answer a research question.
- The focus of this class and the focus of the resources on the Public Health Data LibGuide.
- Uses of secondary data:
 - Disease surveillance, and estimating incidence and prevalence
 - Cross-sectional studies
 - Cohort studies* (what kind?)

Romano.

Secondary Data (advantages)

In general

- Study design and data collection already completed
 - Saves time and money
 - Access to data that would otherwise take several years and millions of dollars to collect
- Can provide population estimates*
- May not need to worry about informed consent, human subjects restriction**

Romano.

Secondary Data (advantages)

In analysis

- Accessible population for primary data may be less representative of the target population than that for secondary data.
- Data may be of higher quality
 - Studies funded by the government generally involve larger samples that are more representative of the target population (greater external validity).*
 - Oversampling of low prevalence groups/behaviors allows for increased statistical precision.**
- Datasets often contain considerable breadth (thousands of variables).

(Koziol & Arthur, nd)

Secondary Data (issues)

In general

- You may need special approval (sign user agreement) to get all variables you need: Name, Address, Telephone, Date of birth, Zip code
- Non-experimental
- Constructs measured by fewer items (no scales) (Shaheen, Pan, & Mukherjee)
- Often requires special techniques to statistically analyze the data
- Missing data are often a problem (Romano)
 - Are data “missing at random” or not?

Romano.

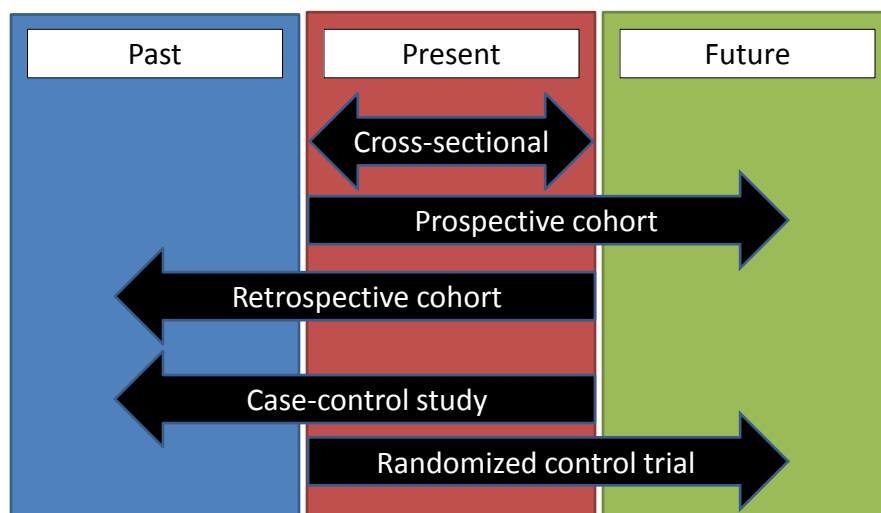
Secondary Data (issues)

In analysis

- Study design and data collection already completed:
 - Data may not facilitate particular research question
 - Information regarding study design and data collection procedures may be scarce
- Data may potentially lack depth (the greater the breadth the harder it is to measure any one construct in depth):
 - Constructs* may be operationally defined by a single survey item which can lead to reliability and validity concerns

(Koziol & Arthur, nd)

Refresher on Study Design



Basic statistical and epidemiology terms used in public health



What is Epidemiology?

- Epidemiology: The study of how a disease is distributed in a population and the factors that influence or determine this distribution (Gordis, 2009).
- The work of an epidemiologist includes (Dicker, Coronado, Denise Koo, & Parrish, 2012):
 - Counting the number of cases or health events, and describing them in terms of time, place, and person;
 - Dividing the number of cases by an appropriate denominator to calculate rates; and
 - Comparing these rates over time or for different groups of people.

Counts

- Counts are whole numbers that represent people, places, and things. Generally, the data is aggregated to protect the confidentiality of participants (Cromley & McLafferty, 2012b).
 - For example, the Census bureau counts to geographic areas such as counties, tracts, block groups and blocks, and reports the aggregate totals.
 - In other cases, data might be summarized using summary statistics such as the average or median.

Densities

- Densities are used primarily in mapping health data.
- A density, such as persons per square kilometer, is a count divided by the area of the geographic unit to which the count was aggregated (e.g., total population divided by number of square kilometers) (DiBiase, Sloan, Baxter, Stroh, & Fletcher King, 2012).

Frequency Measures

- Many variables used in epidemiology are categorical variables, some of which have only two categories (Dicker et al., 2012)
 - Can you think of any examples?
- These variables have to be summarized with frequency measures such as ratios, proportions, and rates.
- Ratios, proportions, and, most important rates are used to describe morbidity (disease), mortality (death), and natality (birth).

Frequency Measures (Ratio)

- Ratio: a comparison of any two values. The numerator and denominator of a ratio do not have to be related (Dicker et al., 2012).
- For example, the sex of children attending an immunization clinic could be compared in either of the following ways:

$$\frac{\text{Number of males}}{\text{Number of females}}$$

$$\frac{\text{Number of females}}{\text{Total number of children}}$$

Frequency Measures (Proportion)

- Proportion: the comparison of a part to the whole. It is a type of ratio in which the numerator is included in the denominator (Dicker et al., 2012).
- You might use a proportion to describe what fraction of clinic patients tested positive for HIV. A proportion may be expressed as a decimal, a fraction, or a percentage.

$$\frac{\text{Number of HIV + clients}}{\text{Total number of clients at your clinic}}$$

Frequency Measures (Rates)

- A rate, is often a proportion, with an added dimension: it measures the number of cases during a certain period of time (usually per year) to the size of the population in which they occurred (Dicker et al., 2012).
- Rates are particularly useful for comparing the frequency of disease in different locations whose populations differ in size (Dicker et al., 2012).

Frequency Measures (Rates)

- The basic formula for rates is:

$$\frac{\text{number of cases or events occurring during a given time period}}{\text{population at risk during the same time period}} = 10^n$$

- What I need to know:
 - Persons in the denominator must reflect the population from which the cases in the numerator arose.
 - Counts in the numerator and denominator should cover the same time period.
 - In theory, the persons in the denominator must be “at risk” for the event.

Frequency Measures (Incidence)

- Incidence, prevalence, and mortality rates are three frequency measures that are used to characterize the occurrence of health events in a population.
- Incidence is the number of (new) cases in a specified period of time in a population at risk for developing the disease (Gordis, 2009).
- Incidence is also known as the **risk**, the **cumulative incidence**, the **incidence proportion** or the **attack rate** for a disease or condition (University of Michigan, 2013).

Frequency Measures (Prevalence)

- Prevalence: The number of (existing) cases out of the number of people at risk or eligible to have the condition (Gordis, 2009).
- The prevalence provides a “snapshot” of the burden of disease at a specific moment in time.
- Prevalence differs from incidence in that prevalence includes all cases, both new and preexisting, in the population at the specified time, whereas incidence is limited to new cases only (Dicker et al., 2012).

Frequency Measures (Mortality)

- A mortality rate is a measure of the frequency of occurrence of death in a defined population during a specified interval.
- When mortality rates are based on vital statistics (e.g., counts of death certificates), the denominator most commonly used is the size of the population at the middle of the time period.
- In the United States, values of 1,000 and 100,000 are both used for 10(n) for most types of mortality rates.

Testing our Knowledge (1)

Table 2.1
Neonatal listeriosis, General Hospital A, Costa Rica, 1989

ID	Sex	Culture Date	Symptom Date	DOB	Delivery Type	Delivery Site	Outcome	Admitting Symptoms
CS	F	6/2	6/2	6/2	vaginal	Del rm	Lived	dyspnea
CT	M	6/8	6/2	6/2	c-section	Oper rm	Lived	fever
WG	F	6/15	6/15	6/8	vaginal	Emer rm	Died	dyspnea
PA	F	6/15	6/12	6/8	vaginal	Del rm	Lived	fever
SA	F	6/15	6/15	6/11	c-section	Oper rm	Lived	pneumonia
HP	F	6/22	6/20	6/14	c-section	Oper rm	Lived	fever
SS	M	6/22	6/21	6/14	vaginal	Del rm	Lived	fever
JB	F	6/22	6/18	6/15	c-section	Oper rm	Lived	fever
BS	M	6/22	6/20	6/15	c-section	Oper rm	Lived	pneumonia
JG	M	6/23	6/19	6/16	forceps	Del rm	Lived	fever
NC	M	7/21	7/21	7/21	vaginal	Del rm	Died	dyspnea

Source: 11

Abbreviations

vaginal = vaginal delivery

Del rm = delivery room

Oper rm = operating room

Emer rm = emergency room

National Institutes of Health
Office of ManagementNIH Library | <http://nihlibrary.nih.gov>

Testing our Knowledge (2)

1. What is the number of Hispanic males living with HIV in Montgomery county as of December 31, 2012?
2. What is the number of persons diagnosed with breast cancer in your community during 2012?
3. I just conducted a study of schizophrenia in Boston, MA. In reviewing state data from 2011, I identified 100 cases of schizophrenia per 100,000 people, living in the greater Boston MSA.

National Institutes of Health
Office of ManagementNIH Library | <http://nihlibrary.nih.gov>

Testing our Knowledge (3)

$$\frac{3,500 \text{ (The number of women in Montgomery Country diagnosed with breast CA)}}{425,000 \text{ (The number of females living in Montgomery County)}}$$

Testing our Knowledge (4)

$$\frac{3,500 \text{ (The number of women in Montgomery Country diagnosed with breast CA)}}{850 \text{ (The number of women in Montgomery county diagnosed with lung CA)}}$$

Major Public Health Data Sources



Demographic Data

- Demography: study of the **static** and **dynamic** aspects of a population.
- Demographic Analysis: study of components of **variation and change** in demographic variables and the **relationships between them**.
- Important aspect of demographics is **civil registration**, which allows us to:
 - Collect data on the vital events happening in a population (live births, deaths, marriages, and divorces).
 - Understand demographic characteristics of different populations at different points in time.

(Becker, 2008)

Census Data

- A census (fancy definition): compiling, analyzing, and disseminating demographic, economic, and social data pertaining to all persons in a country or in a well-delineated part of a country at a specified time (United Nations, 2000).
- A census contains (Becker, 2008):
 - Demographic data (at least age and sex)
 - Economic data (e.g., occupation and income)
 - Social data (e.g., education and housing)

Vital Statistics Data

- Vital Statistics
 - Local governments collect vital record data - data about births and deaths.
- Births
 - Birth records include the mother's residential address, and geographical identifiers.
 - Data about infant births is fairly accurate, but data for the mother can be spotty.
- Death
 - Mortality records generated from death certificates include information about the person and the cause of death.
 - Because of privacy concerns, data is only released in aggregated form (zip code or census tract).

(Cromley & McLafferty, 2012b)

Disease Registry Data

- Disease registries are centralized databases for the collection of information of specific diseases, for example cancer registries.
- Cancer registries are the most extensive disease registries in the U.S..
- National Program of Cancer Registries (NPCR)
 - Established by Congress in 1992, and administered by CDC, the NPCR collects data on the occurrence of cancer; the type, extent, and location of the cancer; and the type of initial treatment (Centers for Disease Control and Prevention, 2013k).

(Cromley & McLafferty, 2012b)

Financial or Service Utilization Data

- Data typically related to direct care at a hospital or clinic.
 - For example, this might include medical provider visits, the number of nights spent in a hospital, or prescription medicine use (O'Hara, 2012).
- Many states also provide utilization data for primary care and specialty clinics, long-term care, home health agencies, and hospices.
- This data is typically linked to an eligibility or enrollment file with demographic data.

CMS BSA Public Use Files

- Basic Stand Alone (BSA) Medicare Claims Public Use Files (PUFs).
- Contains Medicare de-identified claims files, available for public use.
- These files are available to researchers as free downloads in CSV format. They contain non-identifiable claim-specific information and are within the public domain.

CMS LDS Files

- CMS “Non-Identifiable” Limited Data Set (LDS) Files
 - “Beneficiary Encrypted Files,” files that have been stripped of data elements that might permit identification of beneficiaries.
 - Research Data Assistance Center (ResDAC) is a CMS contractor (University of Minnesota) that provides free assistance to academic, government, and non-profit researchers interested in using Medicare and/or Medicaid data for their research.*

CMS OASIS Files

- CMS Outcome and Assessment Information Set (OASIS) files
 - The tool used to collect and report performance data by home health agencies is called the Outcome and Assessment Information Set (OASIS).
 - Since 1999, CMS has required Medicare-certified home health agencies to collect and transmit OASIS data for all adult patients whose care is reimbursed by Medicare and Medicaid with the exception of patients receiving pre- or postnatal services only (Centers for Medicare & Medicaid Services, 2013a).

Other Financial Data Sources

- U.S. Census Bureau Survey of Income and Program Participation (SIPP) data
 - Health Status, Health Insurance, and Medical Services Utilization (2001 to 2011 tables) are available from the Census Bureau Health Data page.
- Veterans Administration Utilization Files
 - The VA Information Research Center (VIREC) develops resources and provides guidance to VA researchers using VA data.
 - Data access is tiered and depends on the use of the data and the data source.

Hospital Data

- Healthcare Cost and Utilization Project (H-CUP)
 - Collection of databases and related software tools and products is made possible by a Federal-State-Industry partnership (AHRQ).
 - HCUPnet provides access to the largest set of all-payer health care databases that are publicly available.
 - Using HCUPnet, you can generate tables and graphs on national and regional statistics and trends for community hospitals in the U.S., and not all states are included.

Survey Data

- Surveys can also be used to screen for particular health conditions. These proactive public health surveys attempt to uncover a health condition before it becomes a problem (Cromley & McLafferty, 2012b).
- The CDC National Health Care Surveys are designed to answer key questions of interest to health care policy makers, public health professionals, and researchers (Centers for Disease Control and Prevention, 2013d).

Major Types of Surveys

- National Surveys: aim to provide data on the entire population under study.
- National (cross-sectional) surveys: involve observations of a population, or a representative subset, at one specific point in time.
- National (longitudinal) surveys: designed to gather information at multiple points in time.

Screening Surveys

- Public health and behavioral surveys are screening surveys that provide information on sub-populations who may be difficult to reach through traditional household surveys.
 - The Behavioral Risk Factor Surveillance System (BRFSS) is the world's largest, on-going telephone health survey system.
 - National HIV Behavioral and System (NHBS) is conducted in rotating annual cycles in three different populations at high risk for HIV.
 - The Youth Risk Behavior Surveillance System (YRBSS) monitors six types of health-risk behaviors that contribute to the leading causes of death and disability among youth and adults.

(Centers for Disease Control and Prevention, 2013)

Keeping Up (1)

RSS at CDC



RSS Feeds

[PODCASTS](#) [PODCAST HELP](#) [RSS](#) [RSS HELP](#)

Stay updated with new content from the CDC. From this page you can subscribe to CDC or other US Government RSS feeds or view their contents directly on this page without having to use an aggregator. New to RSS? See [Podcast Help](#) and [RSS Help](#)

[Features](#) [Browse](#) [Search](#) [Subscribe](#)

CDC RSS Feeds

2010 Haiti Cholera Outbreak	View Items	Subscribe	XML
Attention-deficit/hyperactivity disorder (ADHD)	View Items	Subscribe	XML
Autism	View Items	Subscribe	XML
Birth Defects	View Items	Subscribe	XML
Case Studies in Environmental Medicine	View Items	Subscribe	XML
CDC Clinician Outreach and Communication Activity (COCA)	View Items	Subscribe	XML
CDC E. coli: Brote de actualizaciones en espanol	View Items	Subscribe	XML
CDC E. coli: Outbreak Updates	View Items	Subscribe	XML
CDC Emergency Preparedness and Response: Multistate Fungal Meningitis...	View Items	Subscribe	XML

Keeping Up (2)



Hands-on Exercises

Resources

NIH Library Health Data LibGuide

Douglas Joubert

NIH Library

Testing our Knowledge Exercises

Last Updated:

Exercise 2.2

The line list
General Hos

Medicaid Spending Exercise

Question 1

You have been asked by a patron for data on Medicaid spending (per capita) in Maryland and Pennsylvania.

Background Information

Medicaid

- Medicaid provides health coverage for some low-income people, families and children, pregnant women, the elderly, and people with disabilities. Medicaid programs must follow federal guidelines, but they vary somewhat from state to state.
- In the literature, you will often see Medicaid called a "federal-state" program. This is because the Medicaid program is jointly funded by the federal government and states. The federal

Medicaid Spending Exercise

Exercise 1

You are interested in the approximate per capita on Medicaid in Maryland and Pennsylvania.

Background: Medicaid

- Medicaid provides health coverage for some low-income people, families and children, pregnant women, the elderly, and people with disabilities.
- In the literature, you will often see Medicaid called a “federal-state” program, because Medicaid is jointly funded by the federal government and states.
 - Federal government pays states for a specified percentage of program expenditures, called the Federal Medical Assistance Percentage (FMAP).
 - FMAP varies by state based on criteria such as per capita income. The regular average state FMAP is 57%, but ranges from 50% in wealthier states up to 75% in states with lower per capita incomes (the maximum regular FMAP is 82 %).

(Centers for Medicare & Medicaid Services, 2013)

Background: State Health Facts

- State Health Facts is a project of the Kaiser Family Foundation and provides free access to health data for all 50 states, the District of Columbia, the United States, counties, territories, and other geographies.
- Comprised of more than 800 health indicators and provides users with the ability to map, rank, trend, and download data.
- Data come from a variety of public and private sources, including Kaiser Family Foundation reports, public websites, government surveys and reports, and private organizations.

(The Kaiser Family Foundation, 2013)

What Data Do We Need?

1. ????

What Data Do We Need?

1. Which of the website's databases could you use to find the following information? Statehealthfacts.org
<http://www.statehealthfacts.org/>
2. How much was spent in Maryland?
3. How much was spent in Pennsylvania?
4. What is the approximate per capita expenditure in Maryland compared to Pennsylvania?

FactFinder Exercise

Exercise 2
Where might we get demographic data, other than
State Health Facts?

Background: American Factfinder

- American Factfinder is a search tool that provides access to data about the United States, Puerto Rico, and the Island Areas.
- The following data are available on American FactFinder:

American Community Survey	American Housing Survey	Annual Economic Surveys	Puerto Rico Community Survey	Annual Surveys of Governments
Census of Governments	Decennial Census	Economic Census	EEO Tabulation	Population Estimates Program

(U.S. Census Bureau, 2013)

What Do We Need to Know?

1. ????

What Do We Need to Know?

1. Which type of search should we use?
2. Is this data different (and in which ways) than the State Health Facts data?
3. Are there other methods for interacting with the census data?

Doug Joubert

douglas.joubert@nih.gov